

Weekly Report

06/01/2015 - 06/07/2015

Jing XIA

June 7, 2015

1 Summary

This week I mainly focus on the DataScanner2 project (the second phase of data inspection).

2 Projects

2.1 Project 1 - DataScanner2

DataScanner2: A system that integrates various data quality metrics, covering all quality metrics categories and providing a quality-aware (clustering, correlation, outliers, complex patterns, image quality, feature preservation) visualization.

This week I mainly implemented the web interface and a startup of paper writing, which will be suspended because of the paper rejection. For the web interface part, I sought for a better visual representation and implemented voronoi treemap. The voronoi treemap now can interactive with data exploration.

Reading through the paper review, I feel like our paper was written in a wrong direction. It covers too many paper types: Theory, Technique, Empirical study and System paper, which makes it hard to sufficiently elaborate all. All reviewers thought the similarity metric (mutual information and quartet) idea is good, but the data dissection concept is a distraction. Maybe we should write it more directly. And the reviewers think both cases and the user study are too weak to be convincing.

To be conclude, I think the DataScanner paper should be rewritten with major modifications as below:

The overall concept I think data scanner is OK but data dissection is kind of too obscure to be a good analogy. And maybe it's too early to say "multi-faceted" because there is no non-orthogonal facet yet.

It is still to be decided whether we should cover other quality metrics. I'm afraid it would only make the approach more difficult to be comprehended

and also counter-intuitive. Because I also cannot answer the questions such as “two views are similar in shape, then what”. With two views of high mutual information, analysts can explore where is the high-correspondence region. But with two views of similar shape, analysts can do nothing. Maybe we need other reasonable quality metrics.

Tasks and user study The two are closely related because the exercises in the user study is designed according to the tasks defined. We should alter the task to serve the major features of the similarity metric. The most importantly, how the similarity metric can be utilized to better reveal the dataset.

Cases The same comment applies to all case studies. It shall take more time to do case study, one thing we should have done but did not due to lack of time after fully implementation of the system.

Interactions Once the cases are found, we should find a way to present it better with interactions. Or the other way around, we should try out some explorations to better reveal the “post-similarity” exploration.

Comparison Some reviews said I should compare with traditional methods, or at least state the advantages

Approach elaboration According to the reviews, I also have to elaborate the overall quartet approach better, explaining the technical details better, to make it reproducible.

2.2 Project 2 - NBA Game Visualization

2.3 Project 3 - FlowTracker

3 Paper Reading

Group seminar: GLO-STIX: Graph-Level Operations for Specifying Techniques and Interactive eXploration.

It is an operation-level abstraction of designing/transforming graph layouts. The authors designed and categorized 34 graph-level operations to transform from one layout to another among 6 graph layout such as the force layout, adjacency matrix, PivotGraph, Semantic Substrates, etc. I think it's worth reading because we all need such operation abstraction idea in our exploration design. Especially for the VisComposer project, interactions is too technical or complex to be user-friendly.

4 To Do List

1. Rewriting DataScanner1 whole paper. redo cases and user study.